CENTER FOR THE STUDY OF READING

Technical Report No. 330

THE GROWTH OF CHILDREN'S SIGHT VOCABULARY: A QUICK TEST WITH EDUCATIONAL AND THEORETICAL IMPLICATIONS

Marilyn Jager Adams and A. W. F. Huggins
Bolt Beranek and Newman Inc.

January 1985

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

BBN Report No. 5615
Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

The authors are indebted to Bertram Bruce and John Frederiksen for critical readings of this manuscript, to Brenda J. Starr for her wizardry in all phases of test production and logistics, to Lawrence Zuckerman and Stewart Davis for research assistance, to Alysia Loberfeld for secretarial help with this manuscript, and most of all to the children and staffs of the Cambridge, MA, and Winchester, MA, public school systems for their cooperation in the data collection process. The project was sponsored by the National Institute of Child Health and Human Development (Contract No. N01-HD-7-2836). The production of this manuscript was funded in part by the National Institute of Education (Contract No. 400-81-0030). This paper will appear in Reading Research Quarterly, 1984-1985, Vol. 20.

Abstract

The sight vocabularies of above and below average readers in the second through fifth grades were assessed by having them read a frequency-graduated series of irregularly spelled words with and without context. With or without context the number of correctly read words varied directly and strongly with reading ability. However, when adjusted for the ability to recognize words in isolation, use of context did not appear to vary with overall reading proficiency. Practical applications of the tasks as well as theoretical implications of their results are discussed.

The Growth of Children's Sight Vocabulary: A Quick Test with Educational and Theoretical Implications

In the present study, we compared the abilities of good and poor readers to read a frequency-graduated series of irregularly spelled words presented in isolation and in meaningful context. The experimental tasks were designed as part of a larger effort to develop a test battery for diagnosing difficulties with various word recognition subskills among mainstream students in grades two through five (Adams et al., 1980). As such, their primary purpose was to assess individual differences in children's functional sight vocabularies. Collectively, however, the data also yield useful information on the manner in which the ability to recognize whole words, with and without contextual support, more generally varies with age and overall reading level.

It is repeatedly found that word recognition abilities are the single best class of discriminators between good and poor readers. Moreover, of all the various factors that have been examined--including word-shape, spelling-to-sound correspondences, and numerous correlates of orthographic redundancy--the most powerful determinant of word recognition facility among skilled readers is consistently held to be the readers' familiarity with the words as wholes (e.g., see Adams, 1979, 1981; Broadbent, 1967; Huey, 1908; Johnston, 1978;

McClelland & Rumelhart, 1982; Morton, 1969; Smith, 1971). It follows that a measure of the depth of individuals' sight vocabularies should have considerable practical utility.

A second but equally valuable application of these tasks may be derived from the cross-sectional view they provide of the development of children's sight vocabularies. Although educators are well aware of the importance of adjusting the visual vocabularies of reading materials to their students' level, the question of how best to do so has generally been finessed by recourse to certain standard, "grade-appropriate" word lists. The present tasks offer an escape from the necessarily circular constraint of designing both texts and tests from such lists by providing a means of directly estimating the range of word frequencies that ought to fall within a child's apprehensive capacity.

Finally, the data provide information on the effect of context on word recognition and, in particular, on whether, and if so, how, this effect typically interacts with students' reading ability. This interaction is currently of considerable theoretical interest because schema theory and its kindred interactive processing models (e.g., Adams & Collins, 1979; Perfetti & Lesgold, 1977; Rumelhart 1977, 1980) lead to two opposing hypotheses as to its likely nature (see Adams, 1982). Moreover, because each of these hypotheses leads, in turn, to strong but contradictory educational recommendations, an evaluation of their relative validity is of practical importance as well.

Briefly, schema-theoretic models of reading are based on the constructivist assumption that perception consists in representing or organizing information in terms of one's own, previously acquired knowledge. This assumption is held to be equally applicable at all levels of analysis, from elementary sensory features to complex dimensions of meaning (see Adams & Collins, 1979).

More specifically, it is assumed that the reader's knowledge is organized hierarchically such that the output of any level of processing is the input for the next. In this way the information extracted from the page is, for the mature reader, automatically propagated upward from visual detail through increasingly comprehensive levels of interpretation; this flow of information corresponds to bottom-up processing. Top-down processing occurs as the system searches for information to satisfy partially activated higher level knowledge complexes; for mature readers, this results in automatic priming of the lower level complexes. To oversee these automatic processes, schema-theoretic models have adopted the notion of a central, limited-capacity processor from theories of human information-processing. This central processor is responsible for setting the interpretative goals of the system. The proportion of attentional capacity allocated to higher order dimensions determines whether and how the text will be understood. The proportion allocated to problem areas in the system determines whether and how they will be overcome.

Thus, within schema theory, individual differences in performance may arise in two distinct ways: they may be due to differences either in the reader's relevant knowledge and skills or in the way in which she or he allocates attention to the various subtasks. Depending on how one envisions the interplay between these two factors, one may predict either that good readers should profit more than poor readers from context, or just the opposite.

The first of these predictions follows from the assumption that good readers' relevant knowledge and skills are likely to be more elaborate and more deeply ingrained than are those of poor readers. Because good readers should be more sophisticated with respect to the syntactic and semantic relationships of text, their potential sensitivity to contextual clues should be greater than that of poor readers. Because good readers should be more adept at letter and word recognition, they should also have more processing capacity available for purposes of exploiting contextual clues than should poor readers. In support of this position is widespread evidence that more skilled readers show greater sensitivity to a variety of higher order textual cues (e.g., Cromer, 1970; Meyer, Brandt, & Bluth, 1980; Perfetti & Roth, 1981; Weinstein & Rabinovitch, 1971). Moreover, the hypothesis that good readers' word recognition performance should be more sensitive than poor readers' to context has been a central tenet of the "psycholinguistic" theories of Goodman (1976) and Smith (1971, 1973).

The opposing hypothesis, that poor readers should gain most from context, rests on the recognition that because they are generally such poor decoders, they have the most to gain from context. They can use the syntactic and semantic dimensions of the text as top-down support for their difficult or uncertain bottom-up encoding of the text's visual dimensions. Indeed, by diverting extra attention to the top-down constraints of context, poor readers may often compensate for their decoding difficulties. Conversely, it may be argued that the word recognition performance of good readers is so good without context that there is little room for improvement.

This second hypothesis has been most fully developed by Stanovich (1980) under the title of the "interactive compensatory model," and it too has received considerable empirical support. In particular, studies have shown that the word recognition performance of younger and poorer readers is especially responsive to the presence and compatibility of meaningful context (e.g., Biemiller, 1977-1978; Perfetti, Goldman, & Hogaboam, 1979; Samuels, Begy, & Chen, 1975-1976; Schvaneveldt, Ackerman & Semlear, 1977; Simpson, Lorsbach, & Whitehouse, 1983; West & Stanovich, 1978). Further, semantically appropriate substitution errors are found to be especially frequent among younger and poorer readers (e.g., Allington & Strange, 1977; Biemiller, 1970; Juel, 1980; Kolers, 1975; Weber, 1970), thus lending support to the notion that they are prone to use context to guess the identity of a word instead of worrying over its visual detail.

In short, each of these hypotheses is wholly tenable from a schema-theoretic perspective. Moreover, each is strongly supported by its own contingent of advocates and its own body of experimental evidence. The issue, therefore, is not whether one is correct to the exclusion of the other, but whether it is possible to identify the children or circumstances to which each pertains. This becomes especially important when theoretical implications are translated to educational practice. On the basis of the first hypothesis, it has been suggested that poor readers should be encouraged to depend more on context for purposes of identifying words and discouraged from poring over the words' phonic codes or visual details (see Smith, 1973). In contrast, according to the second hypothesis, poor readers tend to resort to context as a means of compensating for poor decoding skills. In thus circumventing decoding difficulties, poor readers must also circumvent the opportunity to exercise and improve upon the relevant decoding skills. It follows therefore from the second hypothesis that poor readers should be discouraged from relying on context and encouraged to attend to the words' phonic codes and visual detail. In short, the didactic recommendations following from either of these hypotheses are counterproductive from the perspective of the other. In view of this dilemma, a major goal of the experiments to be described was to assess ability-related differences in use of context.

Experiment 1

Even the most meaningful measure is useful only to the extent that it is useable. We were therefore concerned that our method not require cumbersome procedures or laboratory apparatus, but that it be easy to administer and score in the field. The method we ultimately developed involved asking each child to read aloud a list of words of graduated frequency and irregular spelling-to-sound correspondences, such as _island_ and _recipe_. More specifically, the spelling-to-sound correspondences of the test words were not just unusual but at distinct variance from canonical correspondences. This stipulation simultaneously facilitated the scoring procedure and helped ensure that it was the children's sight vocabulary that we were testing (because sounded-out responses were obvious). The words were presented in decreasing order of frequency. We expected that the typical reader would have no trouble with the beginning of the list but would eventually reach a point after which most responses were in error. The point at which this happened was to be our measure of the depth of the child's sight vocabulary.

Method

Subjects. We tested 148 children in the second through fifth grades from an urban public elementary school system in the Boston area. All children were native speakers of English, and none were classified by the schools as dyslexic. Although we found it impossible to equate IQ scores across reading abilities, we excluded children whose Otis-Lennon IQ scores, available from school files, fell below 100 or above 125, to improve the matching of good and poor readers. To verify these scores, we administered the information, vocabulary, picture arrangement and block design subtests of the WISC to each child; if the WISC composite fell below 80 or above 130, the child was dropped from the sample. Stanford and Gates-McGinnitie reading comprehension scores were also obtained for each child. Children who scored within or below the fourth stanine on both tests were classified as poor readers; those who scored within or above the fifth stanine on both tests were classified as good readers. Because some of the standardized test results became available only after our testing had been completed, those students whose stanine scores straddled the above defined boundaries were removed from the sample posteriorly. The final sample included a total of 106 children: 8 poor and 15 good second grade readers, 9 poor and 16 good third grade readers, 16 poor and 13 good fourth grade readers, and 15 poor and 14 good fifth grade readers. The mean age, WISC IQ score, and average reading stanine score is shown in Figure 1 for each group. The children were tested individually in the first semester of the school year.

Stimuli. The stimuli consisted of a list of 50 words with irregular spelling-to-sound correspondences. The frequencies of the words ranged from 134.1 per million to 0.12 per million according to Carroll, Davies, and Richman's (1971) dispersion-adjusted norms (U-scale). The 50 words were selected from a set of 80 words, used in pilot testing with 80 children, so as to exclude words that were of inordinate ease or difficulty given their frequency or that appeared to be beyond the children's listening or speaking vocabularies.[1] The words were listed in decreasing order of frequency as shown in Appendix A.

Procedure. Each child was asked to read the words aloud in order. Children were encouraged to attempt every word on the list. However, testers were instructed that children who seemed especially anxious about their performance and had erred on as many as 10 consecutive words, could be excused from reading the remainder of the list. Responses were scored as "Correct," "Incorrect," or "No Response."

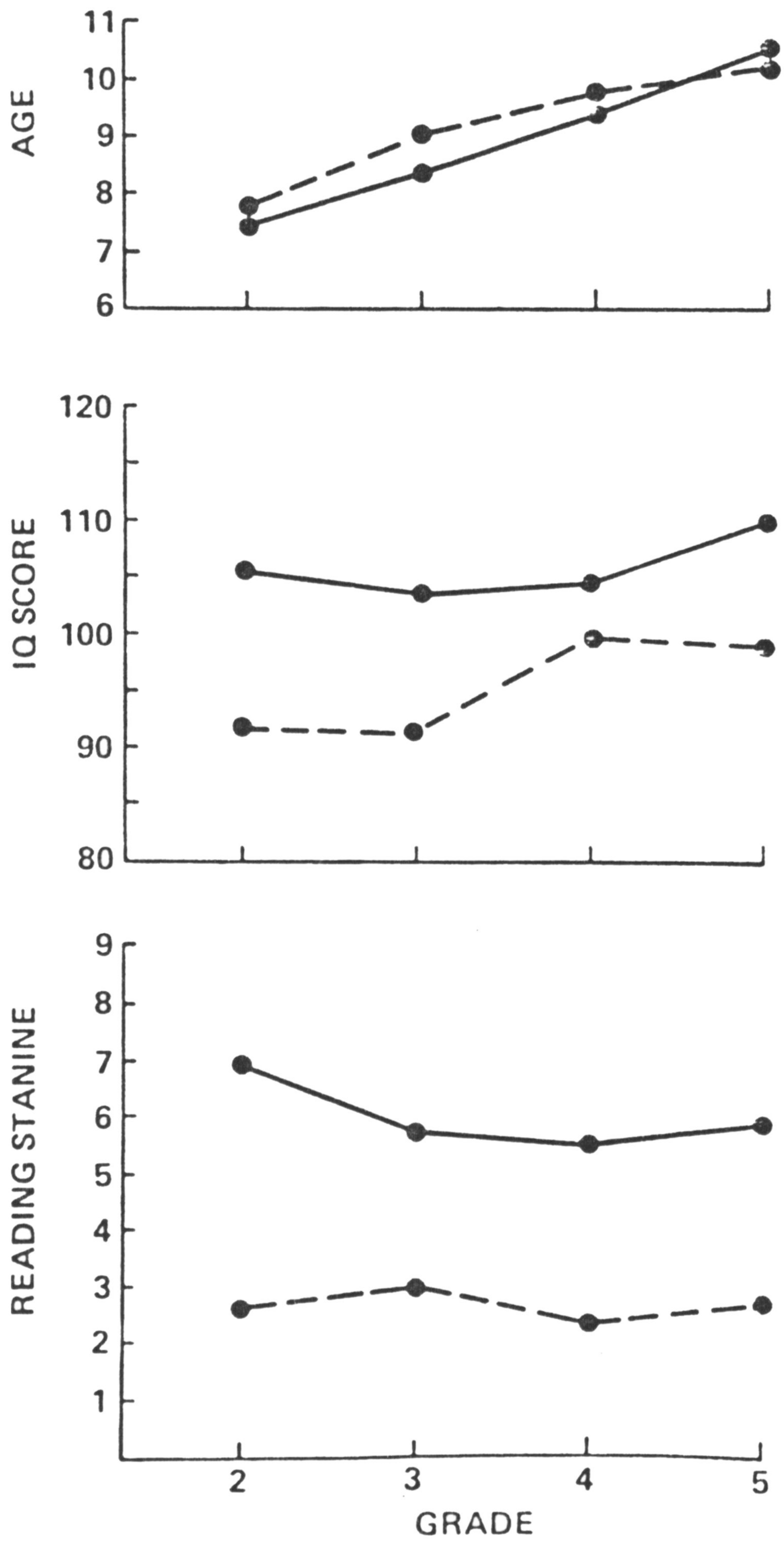

Figure 1. Mean age, WISC IQ score, and reading stanine for good (——) and poor (----) readers in each grade. (Experiments 1 and 2.)

Results

The children's performance very regularly fell into three phases. They would begin reading all words quickly and accurately. Then, for a stretch of 5 or 10 words, their responses would be hesitant and/or occasionally in error. After that point, virtually all responses would be incorrect--typically, the words would be pronounced in accordance with canonical spelling-to-sound translations. Reliability was statistically evaluated through the split-halves method, comparing scores on odd-numbered items with scores on even-numbered items. The correlation coefficient was .932 for the half lists, yielding a reliability of .965 after applying the Spearman-Brown formula to extrapolate back to the full list length.

Mean accuracy for good and poor readers in each grade is shown as a function of decreasing ordinal word frequency in Figure 2. It can be seen from this figure that for all groups, performance declined regularly with decreases in word frequency. Also, the older and better readers penetrated further into the list than the younger and poorer readers. Differences in the number of words correctly read were evaluated through a 4 x 2 (Grade x Reading Ability) analysis of variance. The main effects of both grade, $F(3,98) = 58.41$, $p < .0001$, and reading ability, $F(1,98) = 99.95$, $p < .001$, were highly significant; the

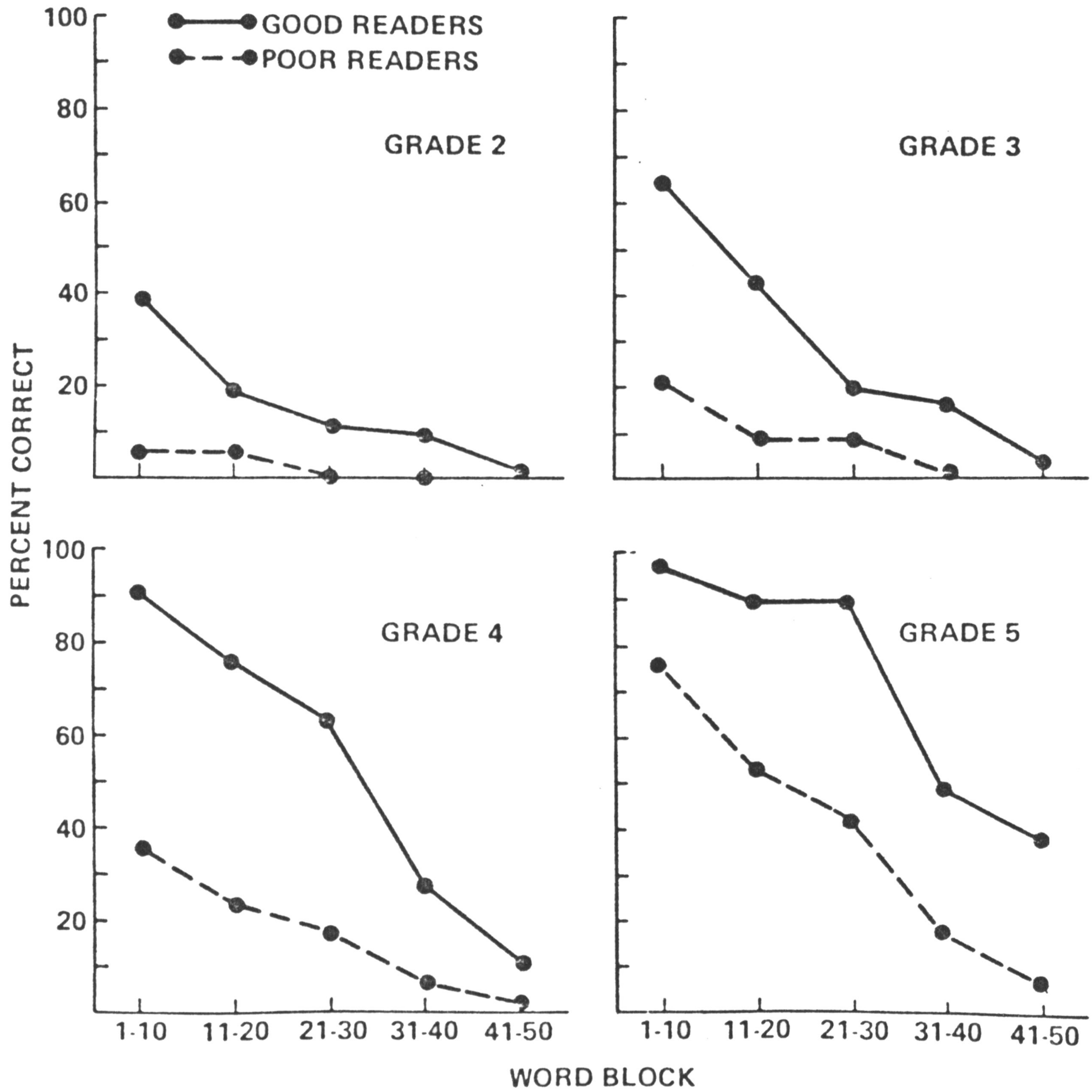

Figure 2. The mean percentage of correctly read words for good (——) and poor (----) readers in each grade. Accuracy is plotted across consecutive blocks of 10 words or, equivalently, as a function of decreasing word frequency. (Experiment 1.)

interaction between grade and reading ability was also significant, $F(3,98) = 5.03$, $p<.01$.

The difference in the mean number of words read correctly by good and poor readers in each grade is presented in Table 1. The interaction we obtained between grade and reading ability is owed to the unevenness of these differences across grades: note, for example, that the difference for the fourth graders is more than twice that for the second graders. The distribution of differences shown in Table 1 might tempt one toward a number of intriguing hypotheses (e.g., about the developmental course of individual differences in reading ability), but it is more likely a relatively uninteresting artifact of our stimulus set. Imagine, for example, two children, both of whom failed to read any words correctly. Suppose that one of these children possessed a stronger sight vocabulary than the other--it is, after all, highly unlikely that the sight vocabularies of any two such children would be identically developed. The point is that even though their scores would have been identically zero on the present list, had the list been extended "backwards" to include more words of higher frequencies, we could have distinguished their different levels of proficiency. By extension, it follows that had the list included more words of higher frequencies, the measured difference between good and poor readers in the lower grades would almost certainly have been greater. Similarly, had the list included more words of lower frequencies, the measured

Table 1

Mean Number of Correctly Read Words and their Differences for Good versus Poor Readers in Each Grade.

	Reading Ability		
Grade	Good	Poor	Difference
2	8.27	1.25	7.02
3	14.69	4.22	10.47
4	26.92	8.31	18.61
5	35.93	19.13	16.80

differences between good and poor readers in the fifth grade might have been larger.

As expected, the number of correctly read words was also strongly related to the children's overall reading ability. The Pearson product-moment correlations between the number of correctly read words and the averaged Stanine scores from the standardized tests were respectively .67, .80, .85, and .82 for the second, third, fourth, and fifth grade children.

Experiment 2

Method

Subjects. The subjects were the same 106 children who participated in Experiment 1. Between Experiments 1 and 2, the children were engaged in a series of five other reading activities which took 20 to 30 minutes.

Stimuli. The 50 test words were the same as in Experiment 1, but each was presented as the last content word of a meaningful sentence. For each sentence, all of the context words were of higher frequency than the test word. The sentences were intended to provide moderate, but not deterministic, priming for the target word. That is, we tried to ensure that several words could be substituted for the test word in each sentence without decreasing the sentence's coherence or likelihood. This was done to minimize the utility of pure guessing: We were interested in children's ability to use contextual information to supplement the orthographic information rather than to substitute for it. The stimuli were again presented in decreasing order of test word frequency. The complete set is shown in Appendix B, with the test words underlined. The test words were not underlined in the list read by the children.

Our decision to use the same rather than different but matched test words in Experiments 1 and 2 was based on pilot testing. The potential problems in using the same words are that the context effect might be inflated because of prior exposure or, conversely, reduced because of perseverative error. Given the goal of assessing individual differences, the problem in using different but "matched" words is that of ensuring that they are indeed "matched" for any given child. The tabulated frequencies of words are, after all, statistical estimates and may be more or less appropriate for any individual. To choose between these alternatives, we constructed two lists of irregular words for pilot testing, List A and List B. There were 40 words on each list and corresponding items were of comparable frequency. Eighty-children served in the pilot test, 20 from each of grades 2 through 5. Half of the children in each grade read the list A words in isolation, and the other half read list B. After 20 minutes of intervening reading activity, all 80 children were asked to read all 80 words, including the 40 they had seen before, in sentential contexts. Across all children, performance on the two lists of words was quite comparable, $t(78) = 0.96$, $p = .34$. We then asked whether the words that had already been seen in isolation were read more accurately in context than those that had not. For the group who read List A in isolation, the answer was marginally positive, $t(39) = 1.40$, $p = .17$; for the group who read List B in isolation, it was slightly negative, $t(39) = -.23$, $p = .82$. A plausible interpretation of these data is that the sentences associated with List A provided stronger contextual cluing than those associated with List B, thereby offsetting what might otherwise have been a consistent advantage to having previewed words in isolation. However, such advantage, if real, was also evidently quite small. Furthermore, Pearson product-moment correlations indicated that performance with the previously seen and unseen words in context was reasonably comparable, with $r(39) = .816$ and $r(39)=.807$ for the respective groups. We therefore decided to use the same words in isolation and in context, on the argument that this option afforded the cleanest interpretation at the level of individual children.

Procedure. The children were asked to read aloud all 50 sentences. As in Experiment 1, the sentences were listed in decreasing order of test word frequency. Testers were instructed to help the children over any difficulties they might have in reading the context sentences but to provide no feedback on the test words. On the children's copy of the list, the test words were not underscored or otherwise set off as special, and the children were not told that we were interested in their reading of but a single word in each sentence. Questioned afterwards, none of the children had recognized the connection between the two experimental tasks.

Results

Performance on the test words was qualitatively similar to that observed in Experiment 1. Again, the number of correctly read words was strongly related to the students' mean reading stanine scores yielding Pearson product moment correlations of 0.86, 0.94, 0.87, and 0.82 for second, third, fourth, and fifth graders, respectively.

The principle difference between the results of the two experiments was quantitative: the children were generally able to read more of the list with context. For purposes of comparison, the mean number of correctly read words both with and without context was evaluated through a 4 x 2 x 2 (Grade x Reading Ability x Experiment) repeated measures analysis of variance, using the unweighted means procedure to correct for unequal group size (Winer, 1971). The effect of experiments or, equivalently, context was highly significant, $F(1,98) = 264.42$, $p < .0001$, as were those of grade, $F(3,98) = 133.44$, $p < .0001$, and reading ability, $F(1,98) = 193.88$, $p < .0001$.

The graphs shown in Figure 3 help to clarify these effects. In particular, note first that, excepting floor and ceiling effects, mean performance for every reader group was consistently superior with context than without and second that, despite this, at any given grade level the mean performance of the poor readers with context never reached that of the good readers without it.

It is worth mentioning that not every child's performance improved with context. The exceptions are listed in Table 2. All but one of these children were second graders, and all but one of them were poor readers. More importantly, virtually all of them read so few words correctly in isolation that their failure to demonstrate improvement with context is uninterpretable: without the reading of a sufficient number of words in isolation, we have no statistically convincing baseline against which to evaluate contextual sensitivity or a lack thereof. It would be wrong to attribute this lack of improvement to an insensitivity to context if, in fact, the very beginning or most frequent portion of our stimulus list was as out of reach for these children as the end of the list was for others. The possible exception is the poor second grade reader who recognized five words in isolation but none in context. This child, however, was indeed a special case as she refused even to attempt the second experimental task and two of the preceding intervening activities.

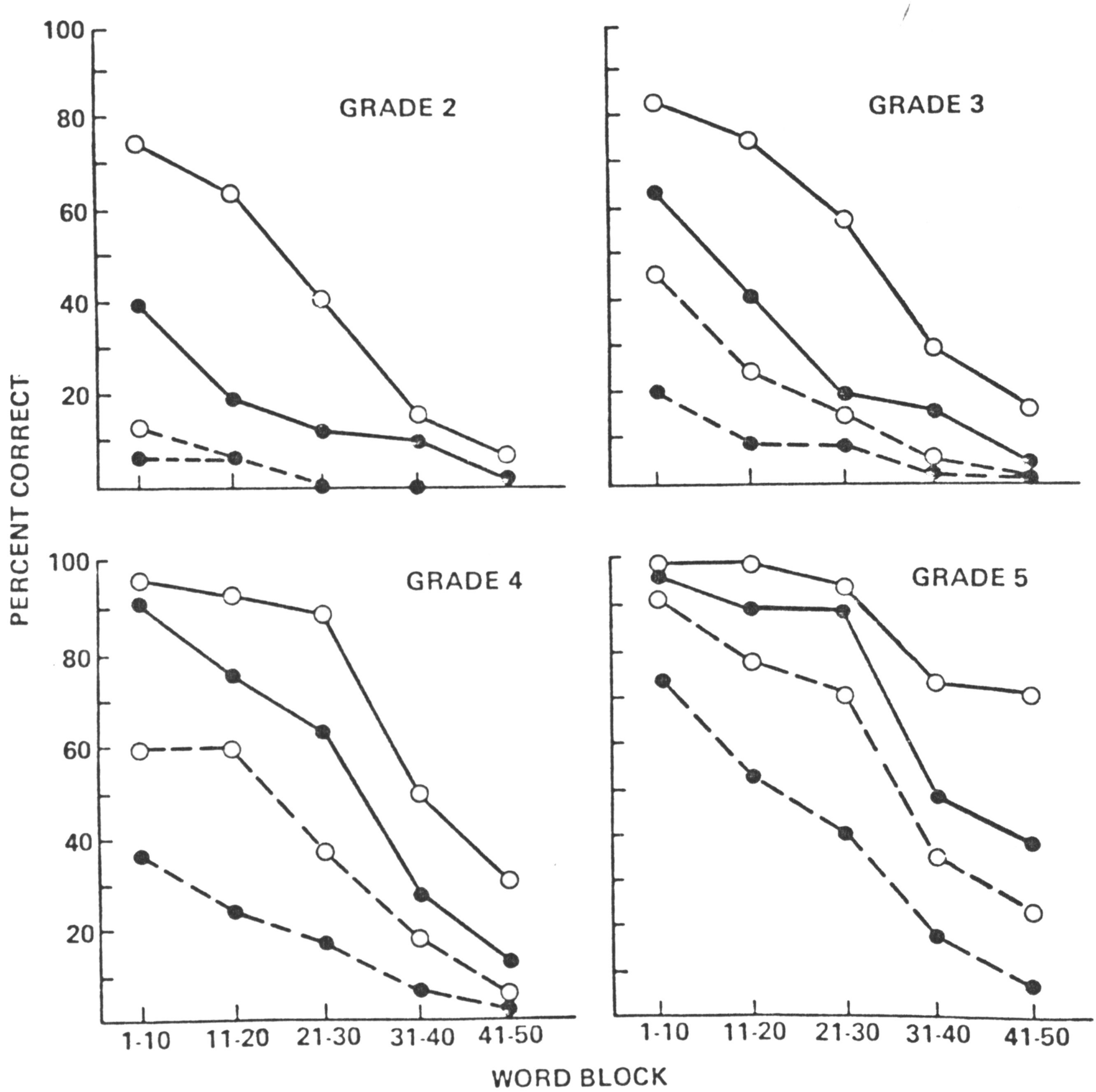

Figure 3. The mean percentage of correctly read words with (○) and without (●) context for good (——) and poor (----) readers in each grade. Accuracy is plotted across consecutive blocks of 1Ø words or, equivalently, as a function of decreasing word frequency. (Experiment 2.)

Table 2

Performance Profiles of Subjects Showing No Improvement with Context

Subjects		Number of Correctly Read Words	
Grade	Ability	No Context	Context
2	poor	0	0
2	poor	0	0
2	poor	0	0
2	poor	1	0
2	poor	1	1
2	poor	5	0
2	good	2	2
3	poor	3	3

Ignoring these subjects, it seems clear that word recognition was significantly aided by the presence of meaningful context for all reader groups in our sample, regardless of age or ability. Yet we may still ask whether some groups were helped more than others. The analysis of variance described above also yielded two significant interactions: the first between experiments and reading ability [$F(1,98) = 7.74$, $p<.01$] and the second between experiments, grade, and reading ability [$F(3,98) = 7.16$, $p<.01$]. While these interactions are not strong enough to challenge the main effects, their significance is consistent with hypotheses that the utility of context is a function of reading ability. An alternative explanation is, of course, that they reflect nothing more interesting than the differential contributions of floor and ceiling effects across reading groups.

To evaluate these possibilities, we directly examined the difference between the two tasks in the percentage of correctly read words. These differences are plotted for each reader group as a function of stimulus words in Figure 4, where word frequency again decreases from left to right. For all of the groups of subjects, these improvement curves exhibit the same, inverted U-shaped characteristic. Relative to the maxima of the curves, the drop in improvement at higher word frequencies (to the left) results from the fact that so many of the words were read accurately without context. The drop at lower frequencies (to the right) indicates that the help that can be gained from

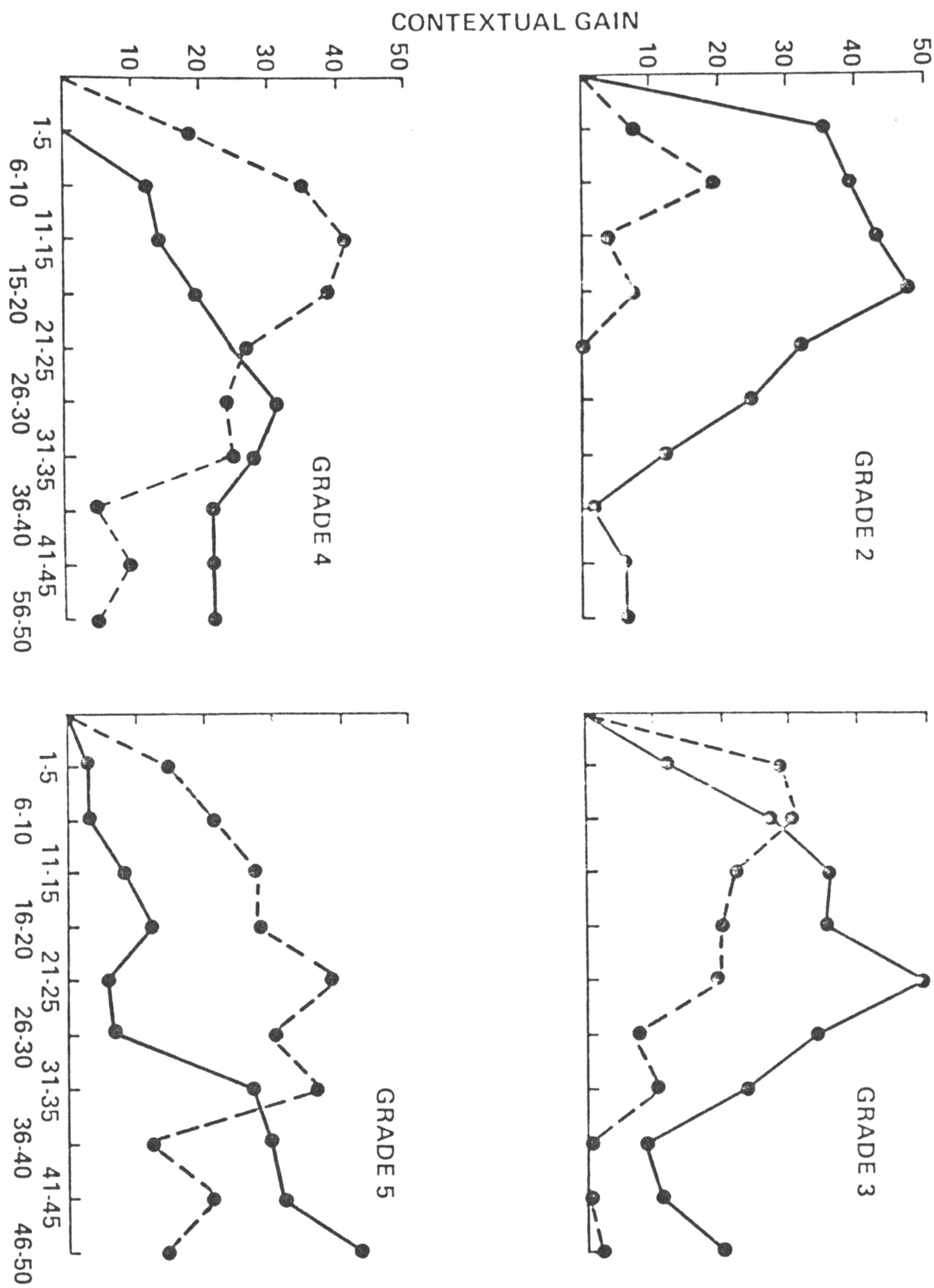

Figure 4. The increase in the percentage of correctly read words when presented in sentential contexts (Experiment 2) rather than in isolation (Experiment 1). The increase for good (——) and poor (----) readers in each grade is plotted over consecutive blocks of five words or, equivalently, as a function of decreasing word frequency.

context is limited; apparently, it cannot totally compensate for a word's lack of visual familiarity. Similar results have been obtained by Frederiksen (1981) and by Pearson and Studt (1975).

The major consistent difference among the curves for the different groups of subjects is in where they reach their maxima. Among the good readers, the vicinity of maximum improvement shifts from a mean word frequency of 24 per million for the second graders to 1 per million for the fifth; among the poor readers, it shifts from a mean frequency of 66 per million for the second graders to 12 per million for the fifth graders. If these peak frequencies are used to gauge the depth of children's sight vocabularies in terms of the number of words acquired, a startling contrast emerges. Using the indices in Carroll Davies and Richman's Word Frequency Book (1971), frequency per million can be translated into rank frequency or, equivalently, into a rough estimate of the total number of words in printed school English that are of higher frequency and that by implication should also be known. For the good and poor second grade readers, these rank frequencies are approximately 2800 and 1200, respectively; for the good and poor fifth grade readers they are approximately 17,000 and 4,500. Once again, these estimates are rough and not only because of the noise in their derivation. It might be argued, on one hand, that they substantially underestimate the children's sight vocabularies because they are based on readings of irregularly spelled words; as regularly spelled words might require less exposure for acquisition, the number of words in the children's more general sight vocabulary might in fact be much larger than these estimates. On the other hand, there is a sense in which these rank frequencies exaggerate the number of usefully distinct vocabulary items learned because the corpus accords separate entries to close morphemic cousins (e.g., shoe/shoes) and typographic variants (e.g., the/The) (see Nagy & Anderson, 1984). But even while we caution against taking the absolute values of these numbers very seriously, we suggest that their relative magnitudes hold important information. Even more sobering than the within grade comparisons between good and poor readers, are the differences between grades in their respective vocabulary growth. The good readers appear to be acquiring well over four times as many new words per year as the poor readers.

Experiments 3 and 4

The results of Experiments 1 and 2 reflect the performance of mainstream students from a relatively low SES, urban population at the beginning of the school year. To test the generality of the response patterns obtained, we replicated the tasks with children from a high SES, suburban school district at the end of the school year.

Method

Subjects. The test was given to 100 students from the second through fifth grade of a suburban public school system in the Boston area. Their Otis-Lennon IQ scores, available from the school files, fell between 85 and 130 points. Reading ability categories were based on Stanford Achievement stanine scores from both the current and the previous school year: Good readers were defined as those scoring within or above the sixth stanine on both reading tests and poor readers as those falling within or below the fifth stanine on both reading tests. (Note that the cutoff was a full stanine higher than for the previous sample; this reflects differences in school norms.) Because some of the Stanford Achievement scores did not become available until after the experiments were run, some subjects (specifically those whose stanine scores straddled our criterion) were eliminated posteriorly. This reduced the analyzed sample to 83 children. These included 10 poor and 9 good second grade readers, 9 poor and 11 good third grade readers, 9 poor and 11 good fourth grade readers, and 12 poor and 12 good fifth grade readers. The mean age, Otis-Lennon IQ and average reading stanine scores are shown in Figure 5 for each group. The children were tested in the last month of the school year.

Stimuli and procedures. The stimuli and procedures were the same as those used in Experiments 1 and 2 except that five test

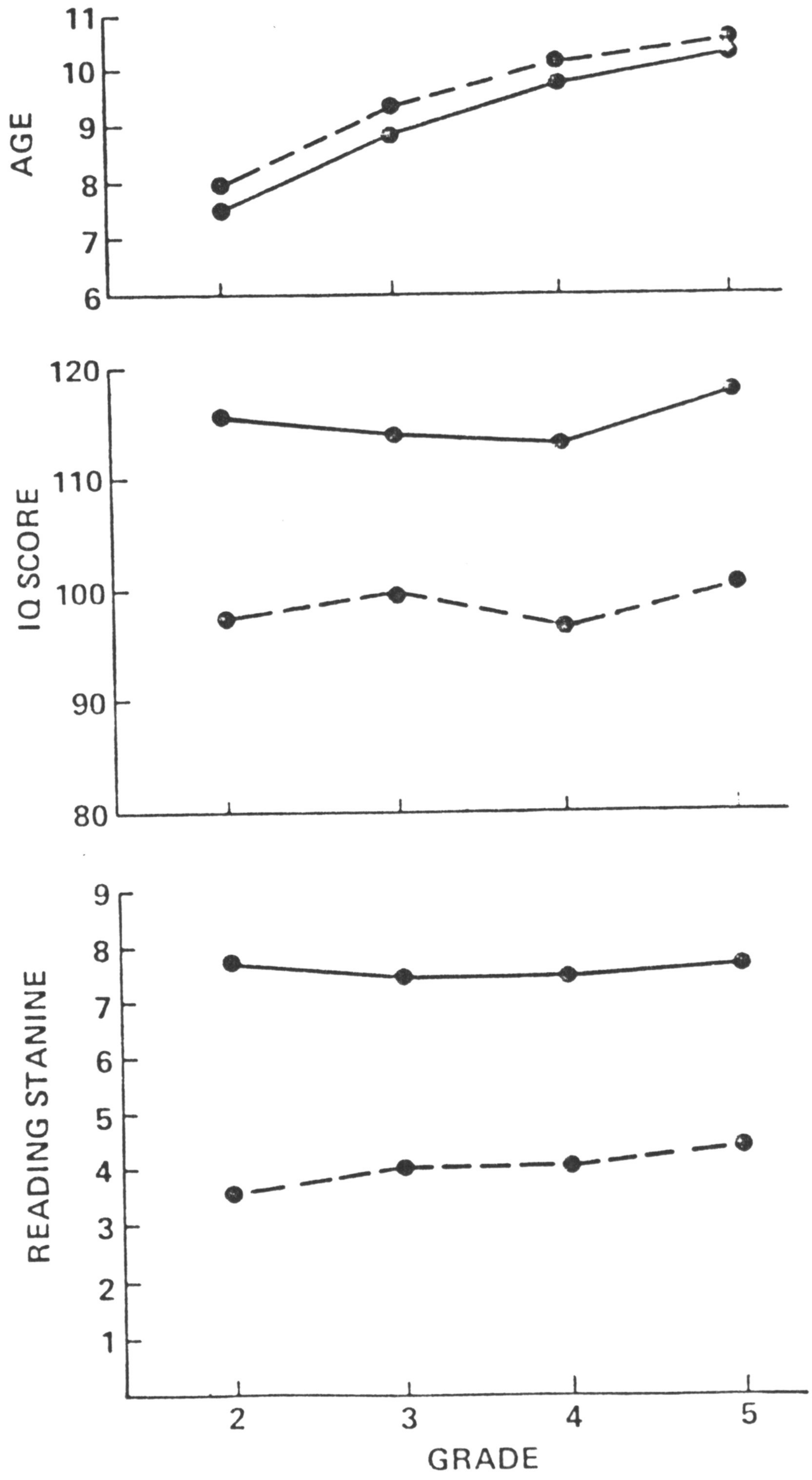

Figure 5. Mean age, Otis-Lennon IQ score, and reading stanine for good (——) and poor (----) readers in each grade. (Experiments 3 and 4.)

words were dropped from the set as part of an effort to reduce the time required to administer the overall battery.[2] The words that were eliminated are marked with an asterisk in Appendices A and B.

Results

Differences in the number of correctly read words were evaluated through a grade by reading ability and experiments (4 x 2 x 2) repeated measures analysis of variance (with corrections as before for unequal group size). Again, highly significant effects of grade [$F(3,74) = 23.86$, $p < .0001$], reading ability [$F(1,74) = 123.94$, $p < .0001$], and experiments (context) [$F(1,74) = 293.83$, $p < .0001$) were confirmed. The interaction between experiments and reading ability [$F(1,74) = 13.17$, $p < .01$] and the triple interaction [$F(3,74) = 4.28$, $p < .01$] were also significant but, as before, are most probably due to floor and ceiling effects.

In short, while quantitatively superior, the results were qualitatively similar to those obtained in Experiments 1 and 2. As can be seen in Figure 6, the performance of all groups improved markedly with context and, with or without context, older and better readers reached further into the test set than younger and poorer readers. It is interesting to note that despite the general superiority of the readers in this sample, the differences between good and poor readers is hardly

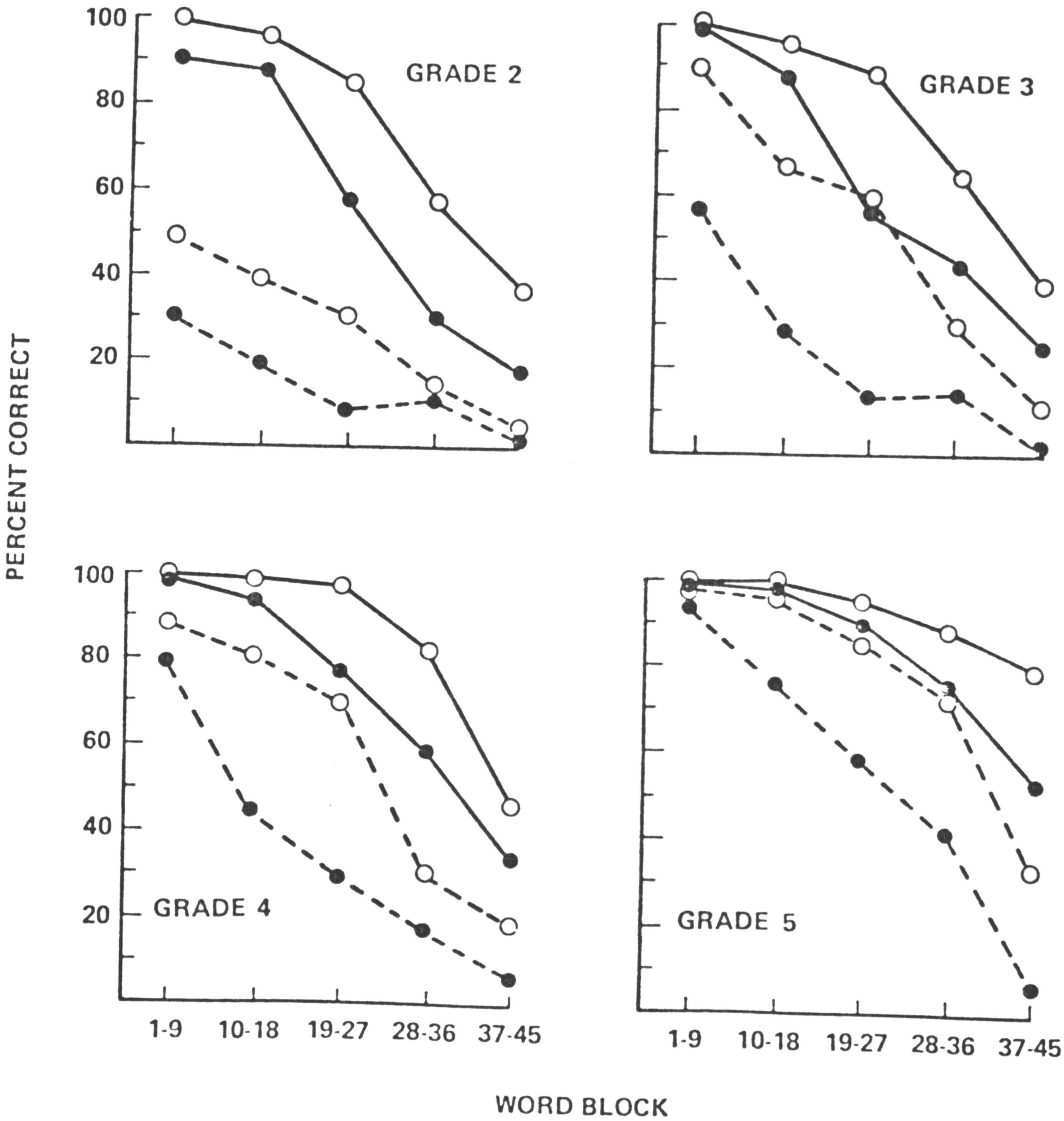

Figure 6. The mean percentage of correctly read words with (O) and without (●) context for good (——) and poor (----) readers in each grade. Accuracy is plotted across consecutive blocks of 9 words or, equivalently, as a function of decreasing word frequency. (Experiments 3 and 4.)

diminished. For example, without context, the average proportions of correct responses were .57, .63, .73, and .83 for second through fifth grade good readers and .14, .24, .37, and .56 for second through fifth grade poor readers: the performance of the poor fifth grade readers did not even meet that of the good second graders. Quite plausibly as a consequence of the general superiority of the reading skills of this group, there was this time only one reader who failed to demonstrate improvement with context: a poor second grade reader who recognized two words correctly in isolation and none in context.

Finally, the correlations between performance on these tasks and the children's reading stanine scores were also quite high whether the words were read in isolation (r = .84, .95, .82, .82 for second through fifth, respectively) or context (r = .84, .85, .84, .79). Floor effects were sufficiently attenuated with this group of readers that the accuracy functions were more comparable and the correlations more equitable across grades than had been the case with our prior sample. In view of this, we decided to extend the correlational analyses one step further. Computing the correlation between the number of correctly read words with and without context, we obtained values of .96, .88, .89, and .92 for the second through fifth graders. There is no hint from these statistics that the relationship of the demands of the two tasks changes with grade level.

Several previous investigators have cited negative correlations between reading ability and contextual gain as evidence for the hypothesis that contextual dependence declines with reading ability. When we evaluated the gain in performance with context against the number of words that were correctly read in isolation, we, too, came up with some sizeable negative correlations: -.404 for all grades combined and .171, -.692, -.572, -.887 for grades 2-5 separately. We feel it is important to point out that although these negative correlations are consistent with the hypothesis that contextual dependence declines with reading ability, they cannot soundly be interpreted as such since the opportunity for improvement also declined regularly with ability. To clarify through example, in the fifth grade, which yielded a high negative correlation, the good readers obtained a mean accuracy of 37.4 words correct in isolation; because the list included only 45 words, they could at best have improved their scores by 7.6 words with context. In contrast, for the second graders, where improvement was not so immediately limited by the impending end of our stimulus list, the correlation was nearly nil. In fact, these examples understate the confounding since subjects with extreme high scores must contribute disproportionately heavily to the correlation statistic even though their performances, within the constraints of our tasks, could vary least.

General Discussion

Contextual Sensitivity and Reading Ability

The results of these experiments unambiguously demonstrate (a) that there are very large differences in the depth of children's sight vocabularies associated both with age or amount of schooling and with reading ability, and (b) that the presence of meaningful context is a potent aid to word recognition regardless of children's age or ability.

However, the data also indicate that the facilitative potential of context is a function of the subjective familiarity of the word to be recognized. Specifically, context assisted recognition most for words of intermediate familiarity: words of greater familiarity were reliably recognized *without* context; words of lesser familiarity were not recognized even *with* context. To be sure, the range of word frequencies that corresponded to "intermediate levels of familiarity" varied directly with age and ability. But the important point is that, for *every* age and ability group, the recognition of words of such intermediate familiarity improved markedly with context.

These results invite speculation on a series of recent studies that focused on the relation between reading ability and contextual sensitivity (Perfetti, Goldman, & Hogaboam, 1979; Schvaneveldt, Ackerman, & Semlear, 1977; Schwantes, Boesl, &

Ritz, 1980; Simpson, Lorsbach, & Whitehouse, 1983; Stanovich, West, & Feeman, 1981; West & Stanovich, 1978; West, Stanovich, Feeman, & Cunningham, 1983). Using lexical decision or target naming tasks, these studies have consistently shown that decreases in response time associated with congruous context vary inversely with reading ability. At first blush, these data would appear to support the second of the hypotheses discussed earlier as they suggest that the extent to which congruous context facilitates word recognition is inversely related to reading ability.

The present data pose both a challenge to this interpretation and an explanation for the data that prompt it. Specifically, in each of the studies just cited, the test words were carefully screened to be within the apprehensive capacity of the youngest or poorest readers in the subject pool. Although the rationale for so selecting the stimuli is obvious, in the context of the present study it can also be seen as a source of significant confounding. That is, if--as the present data attest--the expected familiarity of any given word increases sharply with age and ability, then words that were just familiar to the youngest and poorest readers in these studies should have been quite familiar to the older and better readers. Further, if--again as the present data attest--the facilitative potential of context diminishes as levels of word familiarity increase from marginal to solid, then it must be expected that context would be of less benefit to the older and better readers in these studies.[3] Thus, what might have been interpreted as a decrease in sensitivity to context with ability may more accurately reflect an ability-related increase in the subjective familiarity of the target words.

The present data negate both of our initial hypotheses as to how the utility of context should interact with reading ability: they do not permit us to conclude broadly either that good readers should gain more than poor readers from context or the converse. Instead, Perfetti and Lesgold's (1977, 1978; Lesgold & Perfetti 1978) verbal coding model would seem to present a more useful view of the relation between contextual effects and word recognition abilities. From two premises--(1) that the processes involved in semantic and syntactic integration of text do not differentiate good from poor readers and (2) that the processes involved in word recognition are both the most troublesome and the sine qua non of the reading complex--Perfetti and Lesgold have argued that apparent ability-related differences in sensitivity to context are most often attributable to differences in lexical coding proficiency.

More specifically, Perfetti and Lesgold have argued that when the processes involved in word recognition are especially slow and effortful, they may in themselves press the limitations of the short-term store. In this way, relevant contextual processing is displaced or disrupted such that it cannot influence or be integrated with the incoming text. Perfetti and Lesgold developed this model to explain the relative impoverishment of poor readers' comprehension during on-going reading. However, it provides a plausible explanation of the response pattern reflected by the present data, wherein the interpretive aid of context is lost when the word it has been designed to prime exceeds marginal levels of familiarity.

A Stage Theory of Sight Word Acquisition

On the basis of the present data, we may complement Perfetti and Lesgold's work by suggesting that a word attains sight word status in three stages. At the third or most sophisticated stage, the word is securely represented in the reader's visual lexicon. It is this third stage of mastery that permits the level of word recognition automaticity that is so characteristic of skilled readers and so central to theories of their behavior (e.g., Laberge & Samuels, 1974; McClelland & Rumelhart, 1982). With reference to the present data, only those words that were correctly recognized in the absence of context could be well on their way to being consolidated at this level. Because of the irregular spelling-to-sound correspondences of the words in these experiments, their correct readings in isolation can further be construed as evidence that they were represented per se in their readers' memories. As mentioned earlier, the existence of and a dependence upon whole word representations in memory is also central to most theories of skilled word recognition.

The second or transitional stage of mastery is evidenced by words on our list which were recognized with hesitation or not at all in isolation but correctly with context. Representations of these words have presumably been internalized, but are not yet sufficiently refined or consolidated to support automatic or even reliable direct access. It is therefore words at this stage whose recognition is helped most by the presence of meaningful context. Further, the present data make clear that a given set of words may belong to this stage for one group of readers while having largely migrated to the third stage for a more able group; it is this situation that we hold responsible for the data suggesting that use of context diminishes with reading ability.

Finally, the defining characteristic of words in the first stage of sight acquisition is that they lack any usefully complete internal representation of their orthography. As the balance of the words on our list was not correctly recognized either with context or without, we may infer that most of them belonged properly to this stage of acquisition.

Failures to recognize these words even in context are all the more noteworthy in light of several aspects of the experimental situation. First, the test words were selected under the constraint that they be within the children's listening vocabularies. Although we cannot guarantee that every test word actually met this constraint for every child, we feel secure in asserting that virtually every child's listening vocabulary extended beyond those test words which she or he successfully read aloud. Second, all of the words comprising a sentential context were, by design, of substantially higher frequency than the test word; further, on those rare occasions when a child did display any difficulty in reading the context, the tester provided help. Given that we additionally have evidence that the children were processing the context on earlier sentences, it appears unlikely that failures to recognize these test words coincided with failures to understand their associated context. Third, the children were expected to read to the end of the lists and thus to read well beyond the point where their word recognition accuracy had fallen off. As mentioned previously, a few of the children were excused from reading the entire list of words or sentences, but, at the other extreme, a large proportion of them continued literally to rattle the test words off right or wrong, without hesitation or any other overt sign of difficulty in so doing. Beyond the point of last correct recognition, test words were typically pronounced in solid correspondence with canonical spelling to sound rules. We note that, although distinctly incorrect, such responses can also be seen as not-so-distant approximations to the correct words.

To summarize these three points, test words missed in context were missed despite the facts that (a) they must usually have been well within the child's listening vocabulary, (b) their attendant context was probably interpreted completely and correctly, and (c) their approximations were often decoded with apparent effortlessness to the observer. We may then ask what caused the impasse. Why was the context to no avail? The cause, we suggest, must derive either indirectly or directly from the absence of an internal model of the written word as a whole.

Why Doesn't Context Help with Less Familiar Words?

As a direct explanation for the absence of contextual gain with Stage 1 words, we suggest that perhaps lexical access and the semantic activation it evokes are mediated by and thus depend upon an internal representation of the written word; in the absence of such, processing will be aborted, at best at target naming, but in any case short of the depth of processing necessary to permit contact with the interpreted context. This hypothesis must be qualified in deference to reality constraints. Clearly a complete, previously consolidated word model could not be required for lexical access or it would be impossible to learn to recognize visually new words from reading them which, of course, we do. It therefore seems plausible that on encountering a visually new word, one could create a representation for it using the decoded information and the surrounding context to deduce its lexical identity. For unskilled readers with materials such as those used here, this feat is not likely to be accomplished on a single reading as it should require more or less attention and effort depending on such factors as the amount of effort invested in decoding the word, the proximity of the decoded pronunciation to the correct pronunciation of the word, and the degree of contextual constraint available. The implication, however, is that given multiple readings and directed attention, a number of these test words would eventually have been correctly recognized by our subjects and, as a consequence, entered (probably with Stage 2 status) into their visual vocabularies. This implication also seems plausible to us.

An indirect (but not mutually exclusive) explanation for the unhelpfulness of context with less familiar words can also be offered. Specifically, without the top-down support of an internal representation of the word, the act of decoding may absorb sufficient processing capacity to bump the interpreted context out of working store, thus precluding context/test-word interconnections. In support of this possibility, we refer to some work by Frederiksen (1981). Frederiksen begins with three observations: (a) skilled reading consists in the simultaneous and mutually facilitative execution of a number of information-processing tasks; (b) humans are notoriously limited in their ability to execute multiple information-processing tasks that individually and simultaneously require attention or conscious control; and (c) research on multiple task performance demonstrates that with sufficient practice, processes which at first require devoted attention can become automatic such that they can be performed concurrently without degradation. From these three observations, Frederiksen hypothesizes that the transition to skilled reading occurs only as the component processes develop to the point that their execution is automatic. Prior to this point, their simultaneous and mutually facilitative achievement is precluded by their competing and collectively excessive demands on the reader's active attention.

Using good and poor adult readers as subjects, Frederiksen further conducted a series of experiments designed to assess this hypothesis. From this series, three sets of results are of particular importance to the present discussion. First, Frederiksen obtained vocalization latencies for a set of pseudowords that varied along a variety of orthographic dimensions (lengths, syllabic structures, vowel types, and initial phonemes). The better readers were much quicker at pronouncing the pseudowords than the poorer readers, reflecting a basic difference in the automaticity of raw decoding. Next, Frederiksen compared the pseudoword latencies with those for a carefully matched set of real words. For good readers, the correlations were stronger for low than high frequency words, indicating that the benefit, i.e., the escape from raw decoding, afforded by the presence of the item in the reader's internal lexicon was directly related to the word's frequency or familiarity. In comparison with the good readers, the poor readers' latencies for reading the real word and pseudowords were more highly correlated in general and word frequency made no difference; this indicates that the poor readers generally derived less benefit from the "wordness" of the items and were more dependent on those same raw decoding skills with which they had already demonstrated special difficulty. Frederiksen next presented the words at the ends of meaningful sentences designed to provide strong or weak contextual constraint. For the good readers, the correlations with the readings of isolated pseudowords dropped even more, especially for the high frequency words and the highly constraining contexts; clearly the familiarity of the items and the presence of meaningful context modified their operative constellation of word recognition processes. For the poor readers, however, the correlations remained strong: with the highly constraining context, the correlation showed a slight decrease below that obtained with isolated words; with the weakly constraining contexts, it showed no decrease whatsoever. Again the poor readers' data reflect a heavy dependence on what we have called "raw decoding."

We have reviewed Frederiksen's data as evidence of the breakdown of cooperation between processes that occurs when one or more of them requires extra effort. However, an interesting coda that Frederiksen himself does not raise is that in these cases, it appears to be the more basic process that tends to take precedence. Thus, for Frederiksen's subjects, it appears that the more difficult the task of decoding per se, the less the benefit gained from the familiarity or "wordness" of the item. Further, the less the familiarity or the lower the frequency of the word, the less the benefit gained from contextual constraints on its identity. It is precisely this kind of bottom-up allocation of attention that we are suggesting as an indirect explanation for the absence of contextual gain for the least familiar words in our study.

Extrapolating to Regular Words

It is important to bear in mind that the spelling-to-sound correspondences of the test words used in our study were irregular and that the children's response patterns may accordingly have also been irregular. Indeed, we have evidence from other parts of the same test effort that children's success in reading regularly spelled words aloud depends more on such factors as their orthographic complexity and length than on their frequency (Adams et al., 1980). On the other hand, as this study clearly demonstrates, success in sounding out words quickly and accurately, as per regular spelling-to-sound correspondences, is no guarantee of lexical access or comprehension. If the words were visually novel to the child, then, extending the theory of sight word acquisition proposed above, there is some likelihood that lexical access would not be achieved even if their spelling-to-sound correspondences were regular. Recall that many of the children read erroneously through many of the present test words without hesitation or any other overt sign of difficulty or sensed dissonance. If a child fails to take pause at such necessarily anomalous readings, is there reason to suppose he or she would do so for textually reconcilable ones?

Extending this train of reasoning one more step, we offer the suggestion that the oft-cited phenomenon of word-calling may very often be produced by this very situation. It may reflect competent decoding combined with an effort to keep pace in the face of visually less familiar words. At the very least, if the theory is correct, it underscores the importance of gauging the proximity of the visual vocabulary of a text to its reader's level. In the present study, it was easy to tell when lexical access had failed because the words' pronunciations were then inappropriate. However, had the words been regular, their pronunciations would have been acceptable, and the observer would have been left with no clue that they were not being interpreted.

Summary and Conclusion

Our original purpose in devising these tasks was that of developing a test of children's sight word vocabulary, not a theory of its acquisition. We feel the tasks serve this end quite well: they are quick and easy to administer, strongly discriminative, and straightforward to interpret.

Phenomenologically, the student's behavior with these tasks very much resembles that typically observed with the more familiar reading strand of the Wide Range Achievement Test (WRAT). Indeed, with the broadened perspective of hindsight, we suspect that the major factor controlling performance is one and the same for the present tasks as for the WRAT. The major difference, we suspect, is that because the present tasks involve irregularly spelled words only (and thus preclude access through spelling-to-sound translations), they permit cleaner identification of the point at which direct access falters and, as a consequence, provide a more efficient (shorter) and interpretable test.

Turning to practical applications, the tasks are sufficiently discriminative, and the correlation of their scores with those from the longer, standardized reading tests is sufficiently high, that they might reasonably be used as rough quick estimators of overall reading proficiency. More in line with present interests, by administering both of the present tasks a teacher can estimate the limit of a child's secure sight vocabulary (Stage 3 words) and additionally the boundaries of the child's region of partial acquisition (Stage 2 words). In view of our findings, efforts to stretch a child's vocabulary through independent reading would be best focused on words within this range. In contrast, the acquisition of words beyond this range (Stage 1 words) might be better supported through direct drill and practice or through supervised reading with special efforts to check on comprehension and to enforce rereading wherever potentially troublesome words occur. Thus, while our primary motive was one of developing a tool for assessing the status rather than explaining the course of sight word acquisition, to the extent that our theoretical speculations are correct they bring new dimensions to the importance of having such an assessment tool.

In terms of face-value information, perhaps the most striking aspect of these results is the very marked difference in the sight word vocabularies of above and below average readers. A ready explanation for this difference is that better readers tend to read more text and more sophisticated text so that their opportunity for assimilating new words into their sight vocabulary is greater than that of poorer readers. Surely this is true, but the theory of sight word acquisition presented above suggests an amendment to this explanation. Specifically, the theory suggests that the probability with which a visually novel word will, when encountered, be added to a child's sight vocabulary depends on the child's disposition to attend to the semantic gap otherwise produced and on her or his ability and willingness to invest the necessary thought and effort to close that gap and create a mental token of the word. It is plausible that these metacognitive tendencies and abilities also vary modally with reading ability. However, they would also seem amenable to influence through instruction.

References

Adams, M. J. (1979). Models of word recognition. _Cognitive Psychology_, _11_, 133-176.

Adams, M. J. (1981). What good is orthographic redundancy? In O. J. L. Tzeng & H. Singer (Eds.), _Perception of print: Reading research in experimental psychology_ (pp. 197-221). Hillsdale, NJ: Erlbaum.

Adams, M. J. (1982). Models of reading. In J. F. Le Ny & W. Kintsch (Eds.), _Language and comprehension_ (pp. 193-206). New York: North-Holland Publishing Co.

Adams, M. J. & Collins, A. (1979). A schema-theoretic view of reading. In R. Freedle (Ed.), _New directions in discourse processing_ (pp. 1-22). Norwood, NJ: Ablex.

Adams, M. J., Huggins, A. W. F., Starr, B. J., Rollins, A. M., Zuckerman, L. E., Stevens, K. N., Nickerson, R. S. (1980). _A prototype test of decoding skills_ (Contract No. NØ1-HD-7-2836). Bethesda, MD: National Institute of Child Health and Human Development.

Allington, R. L., & Strange, M. (1977). Effects of grapheme substitutions in connected text upon reading behaviors. _Visual Language_, _11_, 285-297.

Beimiller, A. (1970). The development of the use of graphic and contextual information as children learn to read. Reading Research Quarterly, 6, 75-96.

Biemiller, A. (1977-1978). Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. Reading Research Quarterly, 13, 223-253.

Broadbent, D. E. (1967). Word-frequency effect and response bias. Psychological Review, 74, 1-15.

Carroll, J. B., Davies, P., & Richman, B. (1971). Word frequency book. New York: American Heritage.

Cromer, W. (1970). The difference model: A new explanation of some reading difficulties. Journal of Educational Psychology, 61, 471-483.

Frederiksen, J. (1981). Sources of process interaction in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading (pp. 361-386). Hillsdale, NJ: Erlbaum.

Goodman, K.S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R. Ruddell (Eds.), *Theoretical models an processes of reading* (2nd ed.) (pp. 497-508). Newark, DE: International Reading Association.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press.

Johnston, J. C. (1978). A test of the sophisticated guessing theory of word perception. *Cognitive Psychology*, *10*, 123-153.

Juel C. (1980). Comparison of word identification strategies with varying context, word type, and reader skill. *Reading Research Quarterly*, *15*, 358-375.

Kolers, P. A. (1975). Pattern analyzing disability in poor readers. *Developmental Psychology*, *11*, 282-290.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.

Lesgold, A. M., & Perfetti, C. A. (1978). Interactive processes in reading comprehension. *Discourse Processes*, *1* 323-336.

McClelland, J. L., & Rumelhart, D. E. (1982). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375-407.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. Reading Research Quarterly, 16, 72-103.

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76, 165-178.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English. Reading Research Quarterly, 19, 304-330.

Pearson, P. D., & Studt, A. (1975). Effects of word frequency and contextual richness on children's word identification abilities. Journal of Educational Psychology, 67, 89-95.

Perfetti, C. A., Goldman, S. R., & Hogaboam, T. W. (1979). Reading skill and the identification of words in discourse context. Memory and Cognition, 7, 273-282.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), Cognitive processes in comprehension (pp. 141-183). Hillsdale, NJ: Erlbaum.

Perfetti, C. A., & Lesgold, A. M. (1978). Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. Weaver (Eds.), Theory and Practice in Early Reading (pp. 57-84). Hillsdale, NJ: Erlbaum.

Perfetti, C. A. & Roth, S. (1981). Some of the interactive processes in reading and their role in reading skills. In A. M. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading (pp. 269-297). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), Attention and performance VI (pp. 573-603). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer, (Eds.), Theoretical issues in reading comprehension (pp. 33-58). Hillsdale, NJ: Erlbaum.

Samuels, S. J., Begy, G., & Chen, C. C. (1975-1976). Comparison of word recognition speed and strategies of less skilled and more highly skilled readers. _Reading Research Quarterly_, _11_, 72-86.

Schvaneveldt, R., Ackerman, B. P., & Semlear, T. (1977). The effect of semantic context on children's word recognition. _Child Development_, _48_ 612-616.

Schwantes, F. M., Boesl, S. L., & Ritz, E. G. (1980). Children's use of context in word recognition: A psycholinguistic guessing game. _Child Development_, _51_, 730-736.

Simpson, G. Lorsbach, T., & Whitehouse, D. (1983). Encoding and contextual components of word recognition in good and poor readers. _Journal of Experimental Child Psychology_, _35_, 161-171.

Smith, F. (1971). _Understanding reading_. New York: Holt, Rinehart & Winston.

Smith, F. (1973). _Psycholinguistics and reading_. New York: Holt, Rinehart & Winston.

Stanovich, K. E. (1980) Toward an interactive-compensatory model of individual differences in the development of reading fluency. _Reading Research Quarterly_, _16_, 32-71.

Stanovich, K. E., West, R. F. & Feeman, D. J. (1981). A longitudinal study of sentence context effects in second-grade children: Tests of an interactive-compensatory model. _Journal of Experimental Child Psychology_, _32_, 185-199.

Weber, R. M. (1970). A linguistic analysis of first-grade reading errors. _Reading Research Quarterly_, _5_, 427-451.

Weinstein, P., & Rabinovitch, M. S. (1971). Sentence structure and retention in good and poor readers. _Journal of Educational Psychology_, _62_, 25-30.

West, R. F., & Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. _Child Development_, _49_, 717-727.

West, R. F., Stanovich, K. E., Feeman, D. J., & Cunningham, A. E. (1983). The effect of sentence context on word recognition in second- and sixth-grade children. _Reading Research Quarterly_, _14_, 6-15.

Footnotes

[1] In fact, each of the rejected words proved either aurally unfamiliar or inordinately easy relative to others in its frequency neighborhood. In the latter case, the explanation seemed consistently to be that there were other words with the same root of comparable frequency (e.g., *shoe* and *shoes*).

[2] In retrospect we feel that the lists were slightly better before elimination of these five words. At the time, it was done in the spirit of cooperation to appease other people who were responsible for other parts of the overall test and were asked to trim their contributions.

[3] In contrast to our data, even the most skilled readers in these studies generally demonstrated some facilitative effect of context. This difference can most probably be attributed to the fact that whereas our dependent measure was accuracy, theirs was reaction time.

APPENDIX A

The 50 test words in order of presentation.[a]

1. ocean [b]
2. iron
3. island
4. break
5. busy
6. sugar
7. touch
8. none
9. heights
10. whom
11. tongue
12. lose
13. prove
14. rhythm
15. truth
16. stomach
17. blind
18. wounded
19. calf
20. sweat
21. sword
22. anchor
23. echo
24. guitar
25. veins
26. chorus
27. scent*
28. deaf*
29. mechanic
30. dough
31. rely
32. ninth*
33. react
34. recipe
35. pint
36. deny
37. vague*
38. tomb
39. drought
40. trough
41. depot
42. bough
43. bouquet
44. aisle
45. ache*
46. yacht
47. chauffeur
48. ukelele
49. suede
50. fiance

[a] We would much appreciate hearing from readers who use these lists, with details of how and why they were used, and what results were obtained.

[b] The test words were not numbered on the children's copy of the list.

APPENDIX B

The 50 test sentences in order of presentation.[a]

1. The ship sailed across the ocean.[b]
2. Mary burned her finger on the iron.
3. The girls rowed the boat to the island.
4. If you drop a cup, it might break.
5. Jane could not play because she was too busy.

6. I don't like tea without sugar.
7. The stove is hot, so don't touch it.
8. Ann has two cookies, but Bill has none.
9. He stayed down because he was afraid of heights.
10. I didn't say "what," I said "whom."

11. The hot soup burned her tongue.
12. I like to play games but I hate to lose.
13. She was right but she couldn't prove it.
14. The music was loud and had a good rhythm.
15. The judge asked the man to tell the truth.

16. The football hit him in the stomach.
17. Susan read to the old man because he was blind.
18. The deer was alive but badly wounded.
19. At the farm we saw some pigs and a calf.
20. The hot sun made Joan sweat.

[a] We would much appreciate hearing from readers who use these lists, with details of how and why they were used, and what results were obtained.

[b] The test words were not underscored or numbered on the children's copy of the sentences.

21. The knight killed the dragon with a sword.

22. The crew dropped the ship's anchor.

23. He shouted, and waited to hear the echo.

24. She sang while he played the guitar.

25. Your blood flows through your veins.

26. Sally loved to sing so she joined the chorus.

27. The dogs followed the rabbit's scent.*

28. She didn't hear the bell because she was deaf.*

29. My father took the car to a mechanic.

30. The baker made cookies with the dough.

31. A friend is someone you can rely on.

32. Jeff won the race and Tim came in ninth.*

33. I shouted at him but he didn't react.

34. Father baked the cake from this recipe.

35. We both wanted ice cream so we bought a pint.

36. If you ask mother nicely, she won't deny you.

37. Her memory of what happened was vague.*

38. The hero lay in an unmarked tomb.

39. The corn died during the drought.

40. The horse drank from a trough.

41. The train pulled into the depot.

42. The little bird perched on the bough.

43. The flowers were tied in a pretty bouquet.

44. The pretty girl sat across the aisle.

45. Lifting heavy boxes will make your back ache.*

46. They sailed across the bay on their yacht.

47. The general's car was driven by a chauffeur.

48. He strummed a tune on his ukelele.

49. Her jacket and shoes were both made of suede.

50. She wrote a love letter to her fiance.